LEARNING EVALUATION?

Claire Tourmen
Bear Seminar, 04/08/2014
Berkeley
Why does it matter?

A « near profession »? (Worthen, 1994)

In Education, systematized since the XIXe Century (Foucault, 1975)

« Program Evaluation » since the 1960’s:
« Evaluation is the systematic collection of information about the activities, characteristics, and results of programs to make judgments about the program, improve or further develop program effectiveness, inform decisions about future programming, and/or increase understanding.” (Patton, 2008, p. 39).

• Public organizations and civil servants, NGO’s, private consultants, independent experts in various fields
• + 50 societies in the world (Lavelle & Donaldson, 2009)
• American Evaluation Association, French Evaluation Society
• Accreditation process in Canada since 2010
• Journals and publications
• A majority of training programs in Schools of Education or Psychology in the US (Lavelle & Donaldson, 2009)
French Evaluation Society [www.sfe-asso.fr]

Founded in 1999

Around 300 members:
- 2/3 administrations, civil servants and NGO’s,
- 1/3 evaluators from public or private sectors (researchers, consultants, independent experts)

1 conference / 2 years

Regional clubs and thematic groups of interest

Standards for evaluation, online resources

The Canadian experience

The CES Credential Evaluator Designation program: “The designation means that the holder has provided evidence of the education and experience required by the CES to be a competent evaluator.”

A list of 5 competencies:

1) Reflective practice competencies (i.e. applies professional evaluation standards)
2) Technical practice competencies (i.e. frames evaluation questions)
3) Situational practice competencies (i.e. identifies the interests of all stakeholders)
4) Management practice competencies (i.e. monitors resources)
5) Interpersonal practice competencies (i.e. use negotiation skills)

129 credentialized evaluators in August 2012, 248 in March 2014!

http://www.evaluationcanada.ca/site.cgi?s=5&ss=10&_lang=EN
Publications in the field of program evaluation

*American Journal of Evaluation*

*Foundations of Program Evaluation: Theory of Practice*

*Utilization-Focused Evaluation*

*Evaluation*

Best seller
« 10 little-known fields with great job opportunities »:
Orthoptist, Creative perfumer… and Program evaluator!

« 8. Program evaluator
What you do: You'll evaluate several different programs, making suggestions about changes to make them better, or whether they should even continue. You'll switch programs every few weeks (or whenever you are done evaluating), so you'll get to work with a variety of clients, whether it's a nonprofit, corporate venture or a government initiative.
What you need: A bachelor's degree is sufficient, although some evaluators have a Ph.D. from specialized training programs.
Salary: $56,647”

My research work

• 10 years of research in French, European and international contexts with program evaluators + trainers in vocational training

• A focus on evaluators at work in naturalistic settings: work analysis from psychology of work/development incl. fields observations, clinical interviews (Piaget, 1926) and « explicitation » interviews (Vermersch, 2009)


• How is evaluation performed in real life settings? How is it learned? How can it be taught? Can this knowledge contribute to a theorization of evaluation (Shadish, Cook & Leviton, 1991)?
A piagetian perspective

Jean Piaget’s theory of learning (1947, 1967), as applied to adults’ professional learning at workplace (Billett, 2001, Vergnaud, 2001), that is “The study of how people acquire knowledge” (Mac Carthy Gallagher & Reid, 2007)

To study the development of:

2) the underlying systems of “concepts” and “rules” (or “knowledge”) that structure them in a partially conscious way
A new approach to evaluation

• **Beyond taxonomies of skills** (King et al. 2001, English, 2002)


A developmental approach:

• Before being a method or a discourse, evaluation is a practice

• “Intelligence proceeds from action” (Piaget & Inhelder, 1966/2012, p. 33) and “learning is an internal process of construction” (Mac Carthy Gallagher & Reid, 2007)

• Competencies = capacities to adapt one’s action to the variations of situations
3 research studies

- Program evaluators when designing evaluations at workplace (Tourmen, 2009): 24 practitioners (novices/experienced) in France + European Commission, task analysis, 17 « explicitation » interviews (Vermersch, 2009), 3 field observations, 4 « clinical interviews » based on a case study (Piaget, 1926)

- Experienced program evaluators when describing their past experience (Tourmen, Berriet-Solliec & Lépicier, 2013): 9 in depth « explicitation » interviews with 9 experienced evaluators

- Vocational trainers when evaluating competencies during practical examinations (Tourmen et al., 2013): 1 survey + 10 interviews with trainers, 2 field observations incl. 6 interviews with 2 experienced trainers + 4 students
Trainers evaluating through a practical examination: field observations
Why is evaluation so difficult? Why should it be learnt?

An everyday activity (Stufflebeam, 1980)

Complexity for program evaluators:
• High stakes
• Complex objects
• Strong temporal, technical and budgetary constraints

Complexity for professional trainers:
• Multiple stakes
• Hard to observe
• Limited time and resources

Multiples ways to evaluate and a lot of choices to be made 2 to 10 years to be learnt
I. DIFFERENCES IN ACTIVITY

Programs evaluators
Trainers
Novices were « method-oriented » (Tourmen, 2009)

- **Goals:** “to judge the effects of the program,” “to add knowledge” about it, “to draft well-written evaluative questions,” “to build precise indicators”. “Even if it is an evaluative question, I absolutely don’t know how it could be answered”, “I checked whether I was in line with the definition of the indicator, I was, so I carried on,” “I followed the orders,” “I think that for the formulation of the third evaluative question I didn’t get it right.”.

  = few and short reasoning of anticipation, few goals aimed at and short term goals

- **Situation diagnosis:** « I confess that the methods, I still have trouble to understand which method is going to lead to what, it’s still pretty vague for me », « this is the only thing (indicator) that came to my mind »

  = few and short reasoning of adaptation, few situations indications taken into account to decide, some were neglected
Experienced practitioners were more « results-oriented »

- **Goals**: compromises between several goals and the final results they anticipated they might obtain. They tried “to judge the effects” in a way “that works,” so “it is read” and it “has effects” on the stakeholders. They also tried “to manage without setting the place on fire” in complex actor systems, “to help to sort things out” and “to put one’s finger on the problems.”

  = more and longer anticipation reasoning, more goals aimed at and long term goals

- **Situation diagnosis**: played with the methods according to their diagnosis of the situation, more concerned with the political matters rather than the merely technical ones, more interpretations of the situation. An evaluation manager when building evaluative questions: “The [evaluative] questions will come to the councilors, if I begin to talk about management problems, to say that there are conflicts between some and others… my question might be refused, although it is central.”

  = more and longer reasoning of adaptation, more diverse situation indications taken into account to decide (a broader diagnosis)
Objects to be evaluated
Level of formalization of a program’s objectives
Number, scale, novelty, schedule and level of achievement of a program’s measures
Time of the expected effects
Number of target audiences etc.

People’s strategies concerning evaluation
Origin of the evaluation demands
Explicit/implicit demands
Clarity, scale, compatibility of the demands
Attitudes toward evaluation etc.

The means to proceed with the evaluation
Scale of the resources (time and budget)
Existing information about the program and its effects
Capacity/will of public actors/citizens to participate
Easy/difficult access to informants etc.

CHOICES
Differences in trainers’ activities

As for experienced trainers (Tourmen et al., 2013):

• They did compromises between conflicting goals (assessing the levels with equity, maintaining a high motivation to learn, teaching…), a lot of anticipations of what would happen (« lost of them will be stressed »)

• They were looking at different criteria and indicators while evaluating, even if they were all looking at the same key situation parameters

• They were more or less justifying their assessments depending on the case while debating the final judgments

= Trainers also did a wide diagnosis, aimed at different goals and reasoned to anticipate and adapt
Trainers while evaluating skills: situations indications and parameters diagnosed to make choices (Tourmen et al., 2013)

**Objects to be evaluated**
- Average (if known) skill level of the student
- Success/failure of the students’ activity
- How the student performs activity
- How the student thinks in action, what he probably knows
- Difficulty of the practical examination situation
- Average level of the class
- Examination agenda / teaching
- Etc.

**People’s strategies concerning evaluation**
- Origin and professional project of the student
- Student’s motivation for the subject/examination
- Student’s strategies to gain points
- Contextual/legal requirements for the examination
- Risk of claims/reject of evaluation results
- Student’s level of stress
- Etc.

**The means to proceed with the evaluation**
- Time available
- Practical resources available
- Other jury’s knowledge and evaluation habits
- Evaluation criteria check list
- Etc.
II. DIFFERENCES IN KNOWLEDGE EXPLAIN DIFFERENCES IN ACTIVITY

Program evaluators
Trainers
Certain evaluators reason better than others…

• Longer reasoning
• More conditional reasoning (« If… Then… » statements) using « action rules » (Vergnaud, 2001) to diagnose, predict and adapt
• Taking more goals and situation indications into account

“If I observe this and this and this, then the situation is like this;
If the situation is like this, then I’d better do that;
If I do that, then this will happen”.
Example of expert reasoning

An experienced evaluation manager at a Regional Council building evaluative questions:

“If the evaluative question is too broad, we won’t do it, because, we [the evaluation managers], we begin to measure the consequences in terms of data collection… Automatically we wonder if we have the information available, and if not, can the evaluators have it easily or not…so we restrict the field because we say “the consultants, they can’t make it”…I know what a per diem costs, and if I know that they will have to rebuild a complete data base… I know how much it is. So if I have a €90,000 evaluation and if the consultants will have to rebuild a data base, I will just have one evaluative question, otherwise they can’t sort it out.” (Tourmen, 2009)
... because they have built their own knowledge on evaluation situations!

« Spontaneous theories » (Shadish, Cook & Leviton, 1991) that describe the laws governing the phenomenon

Examples:

- **the way people usually act and react concerning an evaluation**: “you have the managers who want it to be quick, good and clear, you have the civil servants who want evaluation to help them in financing other projects, so sometimes evaluation is an alibi, and you have the influence of consultants who promise you the earth and finally give you undefined stuff.”

- **the way different types of programs usually “happen” and “work”, the way their objectives are decided and their outputs can produce outcomes**: “policies are not always implemented where the stakes are the highest“.

- **the ways different methods can produce different results, their cost and time, their bias and relevance**: “If there aren’t any figures, economists don’t think it’s worth anything. They don’t trust qualitative methods.”
All evaluation practitioners are nascent evaluation theorists” (Shadish, Cook and Leviton, 1991, p. 35.)

Practical knowledge » (Schwandt, 2008, Fitzpatrick, Christie & Mark, 2008, Donaldson & Lipsey, 2008), also called « knowing in practice » (Schwandt, 2005)

This knowledge appeared to be organized around 2 core concepts, relating to 2 main conceptual fields:

- **The feasibility of an evaluation**: the experienced practitioners observed tried to diagnose what would be “feasible,” “measurable,” “possible” or “impossible” and “evaluable” or “hardly evaluable” in a given situation; they also tried to achieve/prevent it with their choices;

- **The legitimacy of an evaluation process and its conclusions**: the experienced practitioners also tried to diagnose and build what would be considered as “sound,” “valid,” “reliable,” “credible” or “robust” conclusions, as opposed to “not serious” and “shaky” conclusions; as well as “useful,” “relevant,” “acceptable” and “independent” conclusions that would not be “rejected.”
Objects to be evaluated
Level of formalization of a program’s objectives
Number, scale, novelty, schedule and level of achievement of a program’s measures
Time of the expected effects
Number of target audiences…
/ compared to the way different types of programs usually « happen » and « work »

People’s strategies concerning evaluation
Origin of the evaluation demands
Explicit/implicit demands
Clarity, scale, compatibility of the demands
Attitudes toward evaluation…
/ compared to the way people usually act and react in this context

The means to proceed with the evaluation
Scale of the resources (time and budget)
Existing information about the program and its effects
Capacity/will of public actors/citizens to participate
Easy/difficult access to informants…
/ compared to the way different methods can produce different results
As for trainers

They also expressed spontaneous theories!

• On students strategies towards evaluation:
  "He managed to give me an answer on another subject to improve his results (...) That’s typical him!"; "He managed his stress much better than I thought"

• On the students’ levels:
  "He says nothing but he’s not... he’s smart", "We all do it" (on a bad practice also done by professionals), "the cow was nice" (on the complexity of the situation)

• On the means to proceed with the evaluation:
  "It is a bit far from real situations... but we still see things" (on the efficiency of the examination)
III. HOW DOES THIS KNOWLEDGE GROW THROUGH EXPERIENCE?

Program evaluators
When does it appear in discourse?

• Not when asked directly! (Tourmen, Berriet-Solliec & Lépicier, 2013)

Evaluator 3:
« (...) 
Question: Is there something you learned about public action during this first evaluation?
Evaluator 3: It don’t understand what you mean?
Question: For instance, did you tell to yourself « things never happen like they are supposed to be » (like just said) or something like that that has appeared to you?
Evaluator 3: Here, no, there is nothing coming back to me, no, I could’nt tell ».

• It comes to mind and words when describing past activities because it is embedded in action (Vermersch, 2009)
• A possible « décalage » between « active thought » and « verbal thought » (Piaget, 1932)
A causal structure of events?

What is it about?

- **Level 1:** assertions about a singular case. Ex: « you could see, on the map, different implementations between what had been done in a department (eq. to county or state) in comparison to others. » (Evaluator 2)

- **Level 2:** generalized assertions relevant for all, many or a type of cases. Ex: « policies are not always implemented where the stakes are the highest » (Evaluator 2)

Forms of “generalizations” (Piaget, 1947) that link event A to event B, and action X to event A or B.

A causal structure linking events to causes (other events/actions) and consequences (other events/actions)?

« People act like social scientists attempting to observe regularities and explain their existence » (Turiel, 1983, p. 4).
An example

Evaluator 2:
- 44 years old,
- Associate Director,
- Studied industrial economics,
- He described the evaluation of agro-environmental measures in 3 regions (seq. to US State levels),
- He expressed a lot of Level 1 and Level 2 theories on a particular phenomenon that struck him, its causes and its consequences.
1. A same policy is implemented in different ways in different aeras
   11. « incoherence of public actors in the implementation of a policy »

2. A director wants to please his Secretary

3. Maybe political pressure of the Secretary’s cabinet

4. The Administration wants to « sell » its measures whatever the conditions
   8. « administrative logic »

5. A disjunction between the measures and the high stake environmental zones

6. Sometimes, vinegrowers were too rich to accept a State aid

7. « policies are not always implemented where the stakes are the highest »

9. « political logic », « it is something usual »

10. « incompatibility problems with the stakes, therefore, relevance problems and coherence problems »

Elements of theory expressed by Evaluator 2
How does it grow?

• A first core idea is developed on a first case (surprise, emotions), then tested, enriched, modified through further experience (« simple idea », Campbell & McGrath, 2011, or « p-prim », Di Sessa, 2000)

• Piaget : “It is during the construction of the rules that learning takes place” (Mac Carthy Gallagher & Reid, 2007) and “experience is central to learning because it enables the child to modify the original theory with which he or she has begun the task” (Mac Carthy Gallagher & Reid, 2007, p. 173)

• Following probabilistic (bayesian) rules of inference? Gopnik & Wellman, 2012: « Intuitive theories are representations of the causal structure of the world (...) in way that goes beyond simple association »

• Role of others and readings in this learning process
Discussion: strengths and limits of this kind of knowledge

<table>
<thead>
<tr>
<th>Process involved</th>
<th>Strengths</th>
<th>Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1 ➔ Level 2</td>
<td>Possibility to:</td>
<td>Risks of:</td>
</tr>
<tr>
<td></td>
<td>• Draw lessons from experience (« learning by doing »)</td>
<td>• Generalize from singular cases</td>
</tr>
<tr>
<td></td>
<td>• Accumulate and share practical knowledge</td>
<td>• Ignore parts of reality that are not directly experienced</td>
</tr>
<tr>
<td>Level 2 ➔ Level 1</td>
<td>Ability to:</td>
<td>Risk of:</td>
</tr>
<tr>
<td></td>
<td>• Anticipate the state of a new program to be evaluated (no time lost, focus on relevant indicators, quicker data interpretation…)</td>
<td>• Confirmation bias (Patton, 2008): tendancy to favor information that confirms our previous beliefs</td>
</tr>
</tbody>
</table>
Synthesis

Evaluators develop practical knowledge - « concepts » and « rules » (Piaget, 1947) - through experience...

...that help them to reason (predict, diagnose and adapt) in a new situation.

= how can we prepare, encourage and secure this learning process among young evaluators?
IV. CONSEQUENCES FOR THE TEACHING OF EVALUATION

Program evaluators
Obstacles to evaluation learning

List of factors (Billett, 2001) relevant for evaluation:
• Long time projects + complex and partly invisible effects = hard to draw lessons from practice
• Variety (confusion?) of methods and guidelines = hard to follow one simple model

“You can be manipulated by different kinds of people, make promises you cannot keep, you have to manage this and anticipate the problems. You can only do it after a lot of setbacks.”
Consequences for evaluation teaching (1)

1) **Need for a first method** to begin, but necessity to go beyond to learn how to read the situations and adapt to them

“If individuals learn that there is just just one way of performing tasks, this might inhibit transfer to other tasks and situations.” (Billett 2001, p. 86).

“The first rules are necessary for gaining initial experiences, but the rules quickly become a barrier to the learning process.” (Flyvbjerg and Sampson, 2001)

2) **Engage in work activities** through guidance by others (Billett, 2001) – in class (case studies) or at workplace
Consequences for evaluation teaching (2)

3) **Vary situations:** “What if you had 50% less budget? What if the program was not finished yet?” - to learn how to face « routine » but also « non routine » and more complex tasks (Billett, 2001), to develop action rules (*if…then…*) and use the core concepts

4) **Reflect on the long term effects of ones’ actions:** work on stories of success and failure to link events - goals – methods – results and help to build action rules/concepts of feasibility/legitimacy
Consequences for evaluation teaching (3)

5) Encourage the « abstraction of principles » or laws and concepts and their discussion (Billett, 2001):

“What kinds of program are the hardest to evaluate? What do people usually think about evaluation? Why do some programs fail to achieve their objectives? Do interviews and case studies provide reliable data?”

Or discuss such polemic assertions as: “qualitative methods fail to provide reliable data; the objectives of a program rarely respond to social needs: nobody reads the evaluation reports” etc.

Possibility to discuss practical principles with scientific concepts/knowledge on programs (political science, sociology, economy, education sciences…) to enrich peoples’ theories and fight against false/limited ones (like “agenda-setting”).

Possibility to write principles on a board as a synthesis of a training
Time for discussion and questions

Thank you for your attention!
References


REFERENCES